This paper not to be cited without prior reference to the author.

A BASIC program for the objective extraction
of growth parameters from length-frequency data

by

D. Pauly[1] and N. David[2]

A new computer-based method is presented for the "tracing" of growth curves through fish length-frequency samples sequentially arranged in time. The method allows for an <u>objective</u> extraction of growth parameters whenever representative samples are available from a fish population. The method is suited both to the analysis of growth in tropical stocks and in temperate stocks, where growth oscillates seasonally. A full program listing with user's instructions is available from the authors.

1) International Center for Living Aquatic Resources Management (ICLARM), MCC P.O. Box 1501, Makati, Metro Manila, Philippines.

2) University of the Philippines Computer Center, Quezon City, Philippines.

Introduction

To the fishery biologist working on stock assessment, a knowledge of how the fish in a given stock grow is essential, since it is generally the growth of individual fish which provide, from year to year, the catch taken by a fishery.

For practical reasons, the available information on the growth of fishes are generally reduced to and expressed by means of a single equation such as the von Bertalanffy Growth Formula (VBGF).

The biological data which can be used to obtain the growth information from which growth parameters can be estimated are of three general types:

- periodic markings (annual or daily) on skeletal parts, scales, otoliths or other bones (or on cartilage in elasmobranchs).
- tagging-recapture data
- size frequency data - most commonly length-frequency data, such as shown in Fig. 1A, 2 and 4.

In spite of frequent criticism, methods for the analysis of length-frequency data have found wider application than both the study of skeletal parts and tagging studies - at least in the case of small, short-lived fishes and invertebrates. The reason for this is that it is usually easier to analyze length-frequency data than to study skeletal parts or conduct tagging studies.

"Paper and Pencil" Methods

The methods presently in use for the analysis of length-frequency data find their origin in the work of Petersen (1892) who pioneered both of the two "paper and pencil" methods still commonly applied to length-frequency samples.
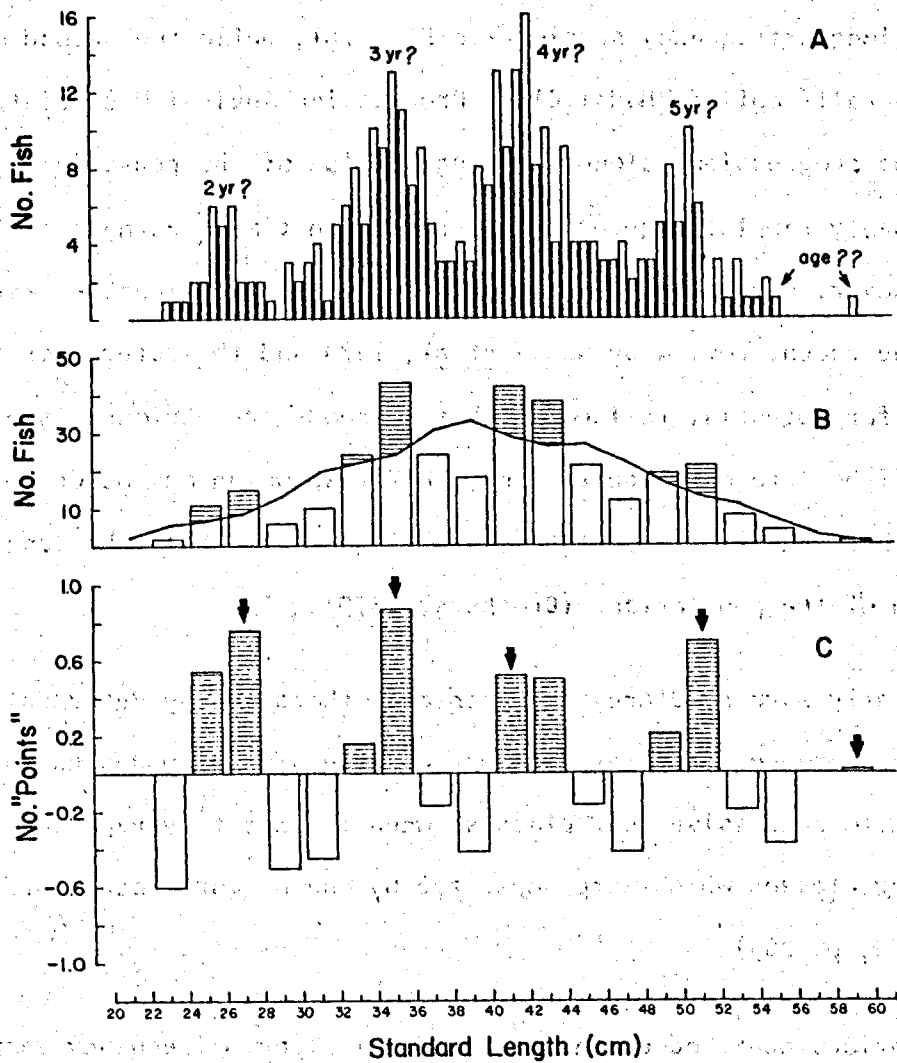
Fig. 1. Length-frequency data on coral trout *(Plectropomus leopardus)* caught near Heron Island (Great Barrier Reef, Australia) in October 1971. From Goeden (1978).

    A. Original data; the ages are from Goeden, with question marks added. N = 319. Note small class interval (5 mm).

    B. Same data, replotted in 2 cm class intervals to smooth out small irregularities, showing running average frefrequencies (over 5 length classes) to emphasize peaks (striped bars above running averages) and intervening troughs.

    C. Same data as in B, after division of each frequency value by the corresponding running average frequency, subtraction of 1 from each of the resulting quotients and subsequent minor adjustments to remove potential sources of bias. Note that "peaks" have been allotted similar numbers of "points," irrespective of the number of fish they represent. Arrows show the "points" used in the computation of USP (see text).

The first of these methods generally referred to as the "Petersen Method", essentially consists of attributing approximate ages to the various "peaks" of a single length-frequency sample (see Fig. 1A), while the second of these methods, generally called "Modal Class Progression Analysis" consists of following the progression, along the length axis, of the peaks in a series of length-frequency samples sequentially arranged in time (Brother, 1979). These methods, however, have often been found to generate questionable results (see e.g. the recent review by Josse et al, 1979 and the references therein). The reasons for potential failures of these methods are obvious: the spawning seasons of fishes are often quite long, and/or spawning may occur in several batches, each of these batches later resulting in a peak in the length-frequency distribution of the population. (Brothers, 1979).

Also,"much early work by D'Arcy Thompson and others, using Petersen's method, was later shown to be inaccurate because a succession of modes had been treated as belonging to successive year classes, when in fact they represented only dominant year classes which were separated by one or more scarce broods" (Ricker, 1975, p. 203).

Thus, the various peaks occurring in any single length-frequency samples often cannot be attributed a definite age, nor unequivocally interconnected with the peaks of preceding, or succeeding samples which results in both the "Petersen Method" and the "Modal Class Progression Analysis" becoming fraught with uncertainties.

Actually, this is not even the worst shortcoming of these methods. An inherent feature of both methods is that the results obtained by different authors from the same set of data generally differ because the methods are essentially based on a subjective interpretation of the data.
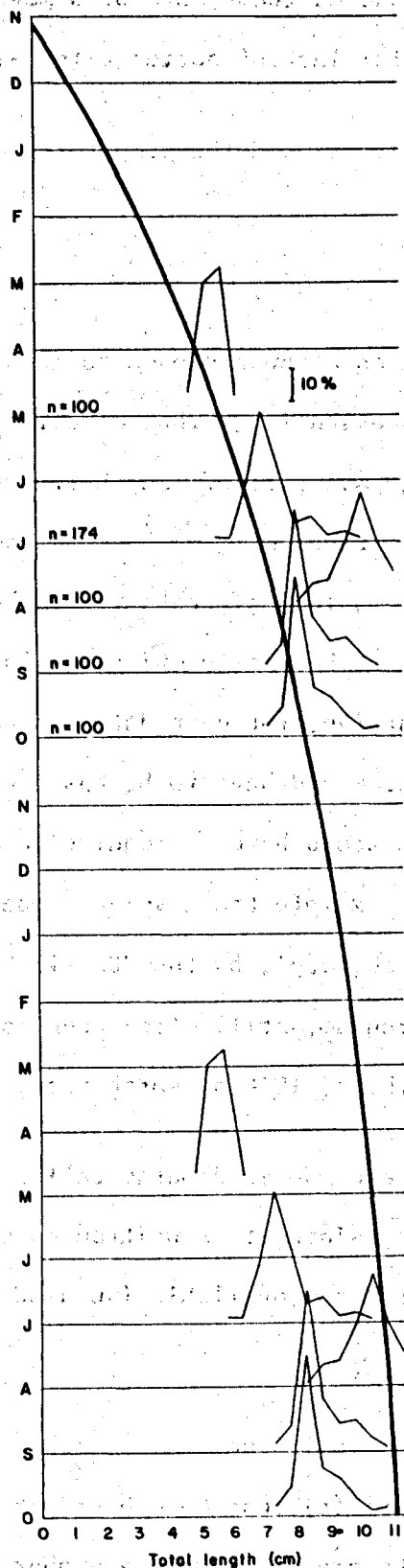
Fig. 2. A set of length-frequency samples arranged sequentially in time, with growth curve fitted by ELEFAN I. Note that the distance between the bases of the samples and the time period between sampling dates are proportional, and that the set of samples is "repeated" one year later, to allow for the forward projection of the growth curve. The curve has, with $D = 1$, the parameters $L_\infty = 12.2$ cm and $K = 1.3$, with ESP/USP $= 0.804$. It must be emphasized that the curve was *not* fitted by eye, and that *no* inputs were made as to expected ages of the various peaks, which of the peaks should be interconnected, etc., (see text). The data, which pertain to slipmouths *(Leiognethus bindus)* caught off Calicut, India, in 1958, were originally published by Balan (1967).

This, more than anything else, is what has rendered these methods somewhat

suspect, even if they are still widely used for lack of better alternatives.

It is with this problem in mind that we have developed the computer program

presented here.

The New Approach

The task we undertook, when starting to work on a new approach to the

interpretation of length-frequency data, was essentially that of devising a

computing procedure that would "trace" through a series of length-frequency

samples sequentially arranged in time, a multitude of growth curves and select

the single curve which, by passing through a maximum of peaks, would "explain"

these peaks. However, the method would have to be wholly objective in the

sense that anyone using the program should arrive, for each data set, at

exactly the same results. The solution, thus would have to be based exclusively

on the length-frequency data themselves, and would have to require no additional

(necessarily subjective) inputs, such as for example the assumed number of age

groups represented in each sample, as required, e.g., by the "Cassie Method"

(Cassie, 1954), or when using the NORMSEP program, still often used for

dissecting length-frequency samples (see Mathews, 1974 or Abrahamson, 1971).

We have succeeded in this, the result being a computer program called ELEFAN I

(Electronic LEngth Frequency ANalysis) which, since it is written in BASIC

can be run on most of the cheap microcomputers now available (we used a Radio-

Shack TSR-80, Level II, 16K).

Put anthromorphically, the program "does" the following:

1) "restructures" the length-frequency sample(s) that have been entered, such

that small but clearly identifiable peaks are attributed a number of "points"

similar to that of peaks based on a larger number of fishes. [The procedure used here essentially consists of calculating running average frequencies (over 5 length classes), dividing each length-frequency value by the corresponding running average frequency, then subtracting 1 from the quotient. A few other minor adjustments are then made to prevent certain type of samples from generating biases.] An example of such restructuring, which is entirely automatic is given in Fig. 1A-C.

2) calculates the maximum sum of points "available"in (a set of) length-frequency sample(s) (see Fig. 1C) ["available points" refers here to points which can possibly be "accumulated" by one single growth curve; see below]. This sum is termed "unexplained sum of peaks" (USP).

3) "traces" through the (set of) length-frequency sample(s) sequentially arranged in time, for any arbitrary "seed" input of L_∞ and K, a series of growth curves started from the base of each of the peaks, and projected backward and forward in time to meet all other samples of the sample set (Fig. 2) and/or the same sample repeated again and again (Fig. 3).

4) accumulates the "points" obtained by each growth curve when passing through peaks (positive points) or through the troughs separating peaks (negative points) (see Fig. 1B and C).

5) selects the curve which, by passing through most peaks and avoiding most throughs best "explains" the peaks in the (set of) sample(s) and therefore accumulates the largest number of points. This new sum is called "explained sum of peaks" (ESP).

6) decrements or increments, the "seeded" values of L_∞ and K until the ratio ESP/USP reaches a maximum, and gives the growth parameters corresponding to this optimum ratio.
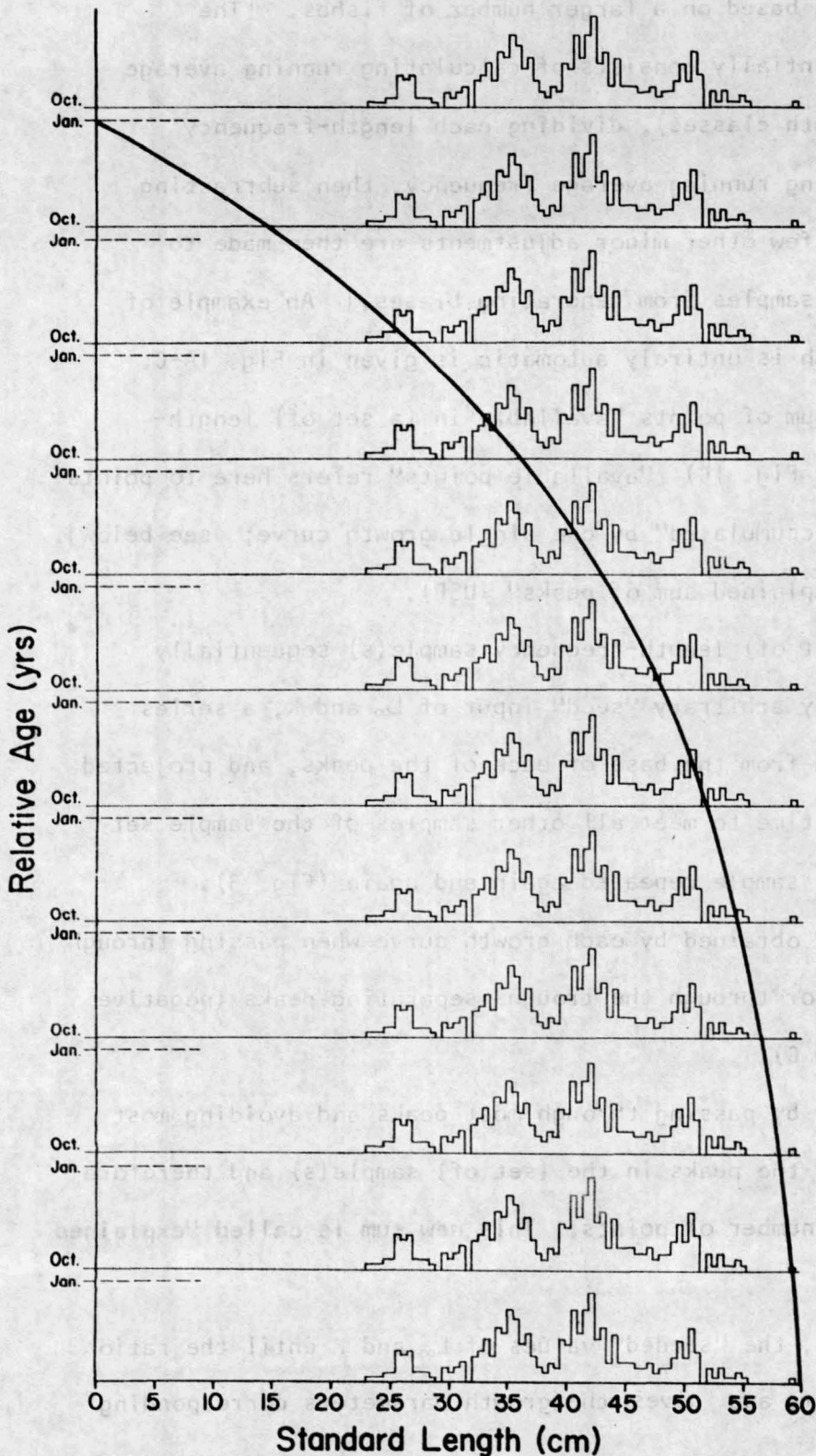
Fig. 3. Length-frequency data on coral trout, fitted with a growth curve by means of ELEFAN I. Note that it is the original sample of Fig. 1A which is shown here, but that the optimization performed by ELEFAN I was based on the "restructured" sample of Fig. 1C. The growth curve has, with D = 1, the parameters L_∞ = 62.4 cm and K = 0.31, with ESP/USP = 0.942. It is again emphasized that the curve was traced without any input except for the length-frequency data themselves.

The curve provides an interpretation of the age structure of the sample different from that originally presented by Goeden. Particularly, what was identified as age group 5 (see Fig. 1A) appears to be age group 6, while the longevity of the fish appears quite higher than originally assumed.

The validity of the procedure outlined here rests on the following assumptions:

1. That the sample(s) used represent the population investigated.

2. That the growth pattern in the population is the same from year to year.

3. That the VBGF describes the average growth of the investigated stock and more specifically:

4. That all fishes in the (set of) sample(s) have the same length at the same age, and that, therefore, differences in length can be attributed to differences in age.

Of these 4 assumptions, the first is a sampling problem and need not be discussed here. Assumptions 2 and 3 appear to be realistic, and they are in fact made - explicitly or not - every time the growth parameters of fish are calculated on the basis of annual markings on skeletal parts.

The last of these assumptions does not strictly apply, since it is known that fishes having exactly the same age may have different length. We feel, however, that no strong bias is generated by making this assumption.

Seasonal growth oscillations

Polar and temperate fishes, and to a lesser extent subtropical and tropical fishes, generally display seasonal growth patterns, with growth being most rapid in summer and slowest in winter, and with fall and spring being seasons of average growth.

In order for our program to be more versatile, and to allow for its use in the analysis of seasonally oscillating length-frequency data, we have therefore incorporated in the program a routine which generates seasonally oscillating growth curves with two additional parameters for expressing the position on the time axis and the intensity of the growth oscillation. Of these two

parameters, one is thus a date of the year, called Winter Point (WP), which corresponds to the time of the year when growth is slowest. The other is a dimensionless constant which expresses the intensity of the growth oscillations and which can take values generally ranging from zero (in tropical fishes) to unity (in temperate fishes) (see Pauly & Gaschütz, 1979, or Gaschütz et al, 1980).

Our "seasonalized" version of the VBGF was derived from a modified version of the standard VBGF which has the form

$$L_t = L_\infty \left(1 - e^{-KD(t-t_0)}\right)^{1/D} \qquad \ldots 1)$$

where D is an additional parameter which can range in fishes between 0.3, such as in large tuna, to 1.0, such as in guppies (Pauly, 1979; Gaschütz et al, 1980). [Setting D = 1, it will be noted, reduces equation 1) to the version of the VBGF commonly used in fish population dynamics.]

The "seasonalized" version of equation 1), has the form

$$L_t = L_\infty \left(1 - e^{-[KD(t-t_0) + C\frac{KD}{2\pi}\sin 2\pi(t-t_s)]}\right)^{1/D} \qquad \ldots 2)$$

where C is the dimensionless constant referred to above, while t_s sets the start of a sinusoid growth oscillation with regards to t = 0; the relationship between t_s and the winter point is given by

$$t_s + 0.5 = WP \qquad \ldots 3)$$

(see Pauly and Gaschütz, 1979, or Gaschütz, et al, 1980).

The seasonality characterizing the growth patterns of many fishes thus generally results in an optimization procedure involving not just 2 (L_∞ and K) but 4 parameters (L_∞, K, WP and C). (See Fig. 4 for an example.)
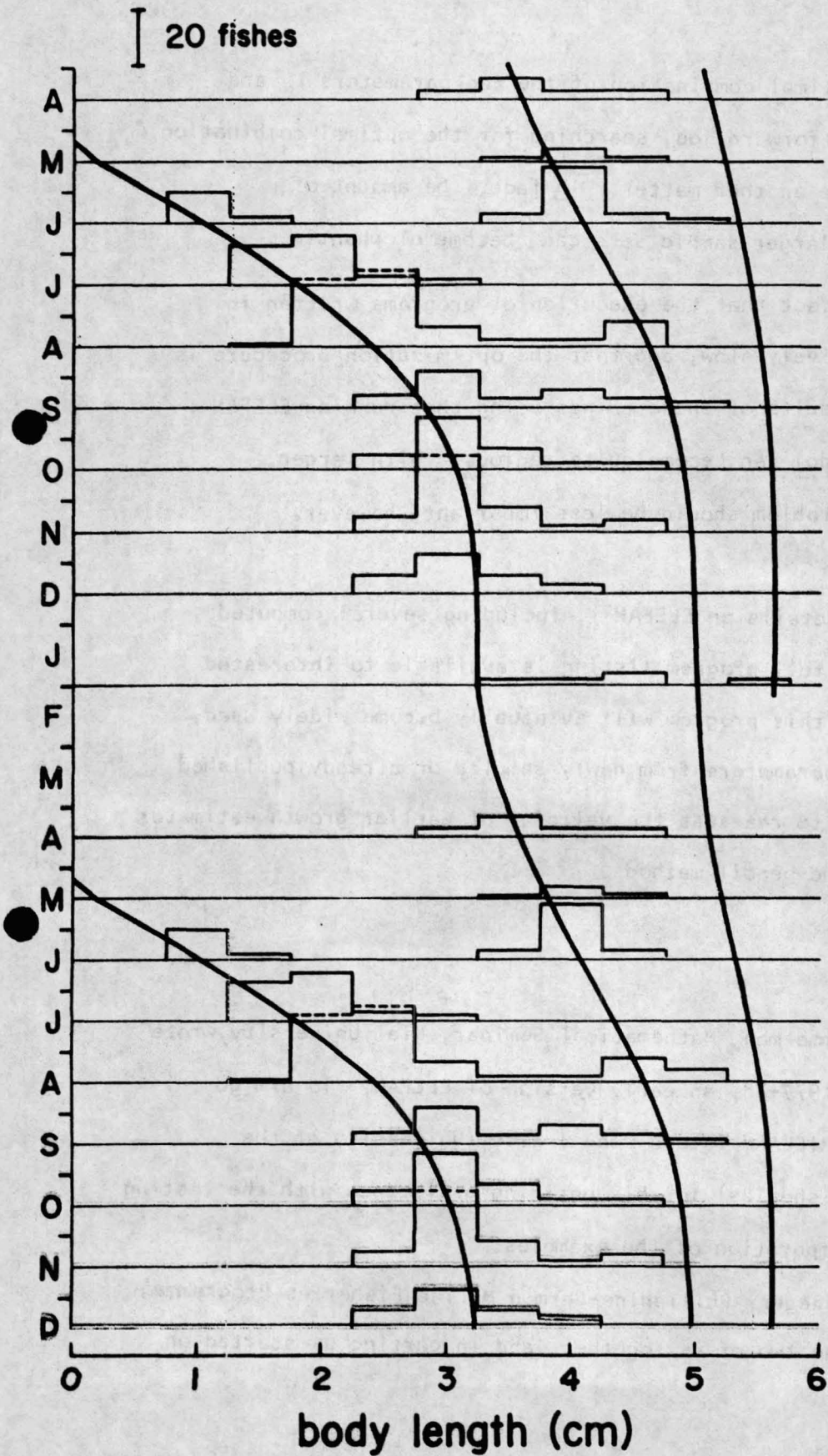
Fig. 4. Length-frequency data on the gobiid *Chasmichthys dolichognathus*, fitted with a seasonally oscillating growth curve by means of ELEFAN I. The original length-frequency data, gathered from April to December 1974 (with the exception of the January-February sample, obtained in 1970) have been here plotted twice to show that the forward projection of the growth curve meets the modal class(es) of most samples, which resulted in the very high value of ESP/USP = 0.933. The growth parameters are, with D = 1: L_∞ = 6 cm, K = 1.0, C = 1.0 and WP = 0, the latter two values corresponding to a brief period of no growth at the turn of the year.

Problems of Computing Time

While searching for the optimal combination of the two parameters L_∞ and

K is a relatively straightforward job, searching for the optimal combination

of four parameters is quite another matter. In fact, the amount of

computation involved with larger sample sets can become elephantine.

This is compounded by the fact that the execution of programs written in

interpreter BASIC is relatively slow, and that the optimization procedure is

partly human-aided, the results of these things being that running ELEFAN I

on a microcomputer (as we do) can become quite tedious. With larger,

faster systems, the time problem should be less important, however.

Availability of ELEFAN I

A report containing more details on ELEFAN I, including several computed

examples and a commented, full program listing is available to interested

colleagues. We hope that this program will eventually become widely used,

both to determine growth parameters from newly sampled or already published

length-frequency data, and to reassess the validity of earlier growth estimates

obtained by using paper-and-pencil methods.

Acknowledgements

Cand. rer. nat. Holger Timmerman, Mathematical Seminar, Kiel University, wrote

for the first author, in 1977-78, an early version of ELEFAN. To him go

our thanks, as well as to ICLARM intern Jose Ingles (University of the

Philippines, College of Fisheries) for his untiring assistance with the testing

of the program and the computation of the examples.

We thank, finally, Dr. J. Saeger, Philippine-German Marine Fisheries Programme,

for his help in getting the two of us together, and in getting us started on

this project.

References

Abrahamson, N.J. 1971. Computer programs for fish stock asssessment. FAO Fish. Tech. Pap. 101. pag. var.

Balan, V. 1967. Biology of the silver-belly Leiognathus bindus (Val.) of the Calicut Coast. Ind. J. Fish. 10(1):118-134.

Brothers, E.B. 1979. Age and Growth Studies on Tropical Fishes. MS presented at the USAID Workshop on Tropical Small Scale Fishery Stock Assessment, University of Rhode Island, September, 1979. (in press).

Cassie, R.M. 1954. Some use of probability paper in the analysis of frequency distributions. Aust. J. Mar. Freshw. Res. 5:513-522.

Goeden, G. 1978. A monograph of the coral trout. Research Bull. No. 1. Queensland Fisheries Services, Brisbane, 42 p.

Gaschütz, G., D. Pauly and N. David. 1980. A versatile BASIC Program for fitting weight and seasonally oscillating length growth data. I.C.E.S. CM 1980/D:6 Statistics Cttee., 23 p.

Josse, E., J.C. Le Guen, R. Kearney, A. Lewis, B. Smith, L. Marec and P.K. Tomlinson. 1979. Growth of skipjack. South Pacific Commission, Occasional Paper No. 11. Noumea, New Caledonia. 83 p.

Mathews, C.P., 1974. An account of some methods of overcoming errors in ageing tropical and subtropical fish populations when the hard tissue growth markings are unreliable and the data sparse. p. 158-166 in: T.B. Bagenal (ed.). Ageing of Fish, Unwin Bros. Ltd., Old Woking, Surrey, England, 234 p.

Pauly, D. 1979. Gill size and temperature as governing factors in fish growth: a generalization on Von Bertalanffy's growth formula. Ber. Inst. f. Meereskunde (Kiel). No. 63, 156 p.

Pauly, D. and G. Gaschütz. 1979. A simple method for fitting oscillating length growth data, with a program for pocket calculators. I.C.E.S. CM 1979/G:24. Demersal Fish Cttee., 26 p.

Petersen, J. 1892. Fiskenbiologiske forhold i Holboek Fjord, 1890-1891. Beretning fra de Danske Biologiske Station for 1890 (91) 1:121-183.

Ricker, W.E. 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. 191, 382 p.

Tamura, E. and Y. Honma. 1977. Histological Changes in the Organs and Tissues of the Gobiid Fishes throughout their Life-Span - VIII Seasonal Changes in Four Species of Gobies. Bull. Jap. Soc. Sci. Fish. 43:893-974.